# 混合分布モデルにおける一致推定量の構成 (区間推定とその関連する問題)



田中, 研太郎

数理解析研究所講究録 1380 巻 2004 年 94-97



# 混合分布モデルにおける一致推定量の構成

東京工業大学　社会理工学研究科 経営工学専攻 田中研太郎
TANAKA Kentaro
The Department of Industrial Engineering and Management
Graduate School of Decision Science and Technology
Tokyo Institute of Technology

**概要**

混合分布モデルは非常に表現力に富んだ汎用性のあるモデルであり, 様々な分野において用いられている. 一方, 混合分布モデルにおいては, パラメータの推定が難しいという問題点があり, 例えば, 最尤推定量が一致性を持たない場合がある事が知られている. そこで, 一致性を持つように制限付けした最尤推定量を構成し, 特に, その制限を標本数の増加とともに緩和可能な推定量の構成を目指す.

## 1 はじめに

混合分布モデルとは, いくつかの確率モデルを組み合わせることによってより複雑な関数形を表現できるようにした確率モデルのことである. 自然現象や社会現象などをモデル化しようとするとき, 母集団が均一でない場合も多く存在し, 結果として非常に複雑な現象が起こっていることが観察され, モデル化が難しいことがある. このような複雑な確率現象のモデル化において, 非常に汎用性の高いモデリング手法を提供する混合分布モデルはとても強力なツールとなる. そして, その高い汎用性から, 混合分布モデルは生物学, 物理学, 社会科学など幅広い分野において用いられている.

一方で混合分布モデルの問題点として, パラメータの推定が困難な場合があることが知られている. とくに, パラメータの推定量としてよく使われる最尤推定量が, 混合分布モデルの場合には必ずしも良い推定量ではなく, それどころか例えば, ロケーションスケール密度関数を成分に持つ混合分布モデルにおいては, 尤度関数が非有界になってしまい, 最尤推定量が計算できなくなってしまう.

実際に混合分布モデルにおいてパラメータを推定する場合には, EM アルゴリズムがよく使われるが, EM アルゴリズムは最尤推定に立脚しており, 実際に EM アルゴリズムを用いてロケーションスケール密度関数を成分に持つ混合分布モデルにおいてパラメータを推定すると, 初期値がうまく選ばれなければ, 尤度関数の非有界性から数値計算が破綻することが確認できる.

本研究では, 制限付きの最尤推定量を扱うことによってパラメータ推定における問題を回避できる事を数理的に裏付けた.

## 2 混合分布

位置を表すロケーションパラメータと，尺度を表すスケールパラメータを持つ密度関数をロケーションスケール密度関数という．正規分布は，平均をロケーションパラメータとし，標準偏差をスケールパラメータとして持つロケーションスケール密度関数である．

$M$ 個のロケーションスケール密度関数を成分に持つ混合分布の密度関数を

$$f(x;\theta)=\sum_{m=1}^{M}\alpha_m f_m(x;a_m,b_m)$$

と表す．ここで，$a_m$ はロケーションパラメータで $b_m$ はスケールパラメータであり，$\alpha_m$ は重みを表す．パラメータ空間 $\Theta$ は

$$\Theta=\{\theta=(\alpha_1,a_1,b_1,\ldots,\alpha_M,a_M,b_M)\in\mathbb{R}^{3M} \\ \mid 0\le\alpha_1,\ldots,\alpha_M\le 1\ ,\ \sum_{m=1}^{M}\alpha_m=1\ ,\ b_1,\ldots,b_M>0\}$$

であるとする．パラメータ空間はユークリッド空間の部分集合であるとし，2点 $\theta,\theta'\in\Theta$ の距離を $\mathrm{dist}(\theta,\theta')$ で表すことにする．

## 3 一致推定量の構成

良い推定量の基準として強一致性があり，それは以下で定義される．

**定義 3.1.** (強一致性) 真の分布を表すパラメータ全体を

$$T\equiv\{\theta\in\Theta\mid f(x;\theta)=f(x;\theta_0)\quad a.e.\ x\}$$

と書くことにする．ここで，$\theta_0$ は真の分布を表すパラメータのうちの1つである．推定量 $\hat{\theta}_n$ が以下の式を満たすとき，その推定量は強一致性を持つという．

$$\mathrm{Prob}\left(\lim_{n\to\infty}\inf_{\theta'\in T}\mathrm{dist}(\hat{\theta}_n,\theta')=0\right)=1$$

つまり，確率1で真の値に近づく推定量のことを強一致性を持つという．

$n$ 個の標本 $x_1,\ldots,x_n$ が得られたとき，尤度関数 $\prod_{i=1}^{n}f(x_i;\theta)$ を最大にするパラメータ $\theta$ を最尤推定量という．ロケーションスケール密度関数を成分に持つ混合分布においては，ある成分のロケーションパラメータをある標本の値と等しくとり，スケールパラメータを0に近づけると，尤度関数が無限大に発散し，最尤推定量が強一致性を持たない．本研究では制限付きの最尤推定量を考え，これが強一致性を持つことを示した．

標本数 $n$ の増加とともに広がっていくパラメータ空間 $\Theta_n$ を

$$\Theta_n = \{\theta \in \Theta \mid 0 < c_n \leq b_m \ , \ m = 1, \ldots, M\} \ .$$

とする．また，以下の正則条件を課す．

**Assumption 1.**
ある実数 $v_0, v_1 > 0$ と $\beta > 1$ が存在して，

$$f_m(x; a_m = 0, b_m = 1) \leq \min\{v_0 \ , \ v_1 \cdot |x|^{-\beta}\}$$

をすべての $m$ について満たす．

これは，$f_m$ $(m = 1, \ldots, M)$ が有界で裾が $|x|^{-\beta}$ よりはやく減衰することを意味する．$\Theta$ の任意のコンパクト部分集合を $\Gamma$ で表すとする．

**Assumption 2.** 任意の $\theta \in \Theta$ と任意の正の実数 $r$ に対して，

$$f(x; \theta, r) \ \equiv \sup_{\mathrm{dist}(\theta', \theta) \leq r} f(x; \theta').$$

とおいたとき，各々の点 $\theta \in \Gamma$ と十分小さな $r$ に対して，$f(x; \theta, r)$ は可測．

**Assumption 3.** $\theta \in \Gamma$ に対して，もし，$\lim_{n \to \infty} \theta_n = \theta$ なら，

$$\lim_{n \to \infty} f(x; \theta_n) = f(x; \theta)$$

となる．（列 $\{\theta_n\}_{n=1}^{\infty}$ に依存しない零集合以外で収束．）

**Assumption 4.**

$$\int |\log f(x; \theta_0)| \, f(x; \theta_0) \mathrm{d}x < \infty.$$

**定理 3.2.** $E_0[\cdot]$ で真の分布による期待値を表すとする．$\mathscr{G}_K$ を $K$ 個の成分を持つ *subprobability measure* の集合とする．

$$\mathscr{G}_K \ \equiv \ \{\sum_{m=1}^{K} \alpha_m f_m(x; \eta_m) \mid \sum_{m=1}^{K} \alpha_m \leq 1 \ , \ \alpha_m \geq 0\}$$

Assumption 1–4 が満たされていると仮定する．また，真のモデルは $M$ 個の成分を持つモデルのみによって表されるとする．このとき，ある実数 $\lambda, \kappa > 0$ が存在して，

$$E_0 \left[\log \{g + \kappa\}\right] + \lambda < E_0[\log f(x; \theta_0)]$$

が全ての $g \in \mathscr{G}_L$ ，$(L \leq M - 1)$ に対して満たされる．

**定理 3.3.** Assumption 1–4 が満たされていると仮定する. $M$ 成分からなる有限混合分布の真の密度関数 $f(x;\theta_0)$ が, $(M-1)$ 以下の成分では表せないとし, ある実数 $u_0, u_1 > 0$ と $\beta > 1$ が存在して,

$$f(x;\theta_0) \leq \min\{u_0\ ,\ u_1 \cdot |x|^{-\beta}\}$$

を満たすとする. $c_0$ を正の実定数とする. そして $\eta$ を $0 < \eta < 1$ を満たす正の実定数とする. 全ての $n$ に対して $c_n = c_0 \cdot \exp\left(-n^{(1-\eta)}\right)$ であるとき, $\Theta_n$ における最尤推定量は強一致性を持つ.

# 4 数値実験

$g(x;a,b)$ で区間 $[a-b, a+b)$ 上の一様分布の密度関数を表すとする. 真の密度関数を

$$0.6 \cdot g(x;0.5,0.5) + 0.4 \cdot g(x;0.6,0.2)$$

としたときに, モデルとして

$$0.6 \cdot g(x;0.5,0.5) + 0.4 \cdot g(x;a,b=c_n), \quad c_n = \exp(n^{-0.95})$$

を考える. モデルのパラメータは $a$ のみである. 標本数とそれぞれの場合の対数尤度 (尤度関数の対数) とを数値計算して表 1 の結果を得た. 標本数の増加とともに, 真の密度関数における対数尤度の値が, モデルに対して優越していくことが分かる.

表 1: 真の密度関数とモデルにおける対数尤度

| 標本数 $n$ | 対数尤度 (真) | 対数尤度 (モデル) |
|---|---|---|
| $10^1$ | 1.757549 | 2.706045 |
| $10^2$ | 10.70968 | 10.70968 |
| $10^3$ | 114.9434 | 196.0215 |
| $10^4$ | 1117.067 | 1200.219 |
| $10^5$ | 11357.03 | 5150.472 |
| $10^6$ | 116656.9 | -9639.489 |